# Package 'ECDFniche'

April 27, 2026

**Type** Package

**Title** Empirical Cumulative Distribution Function Niche Modeling Tools

**Version** 0.5

**Description** Simulate ecological niche models using Mahalanobis distance, transform distances to suitability with 1 - empirical cumulative distribution function and 1 - chi-squared, and generate comparison figures.

**License** MIT + file LICENSE

**Encoding** UTF-8

**RoxygenNote** 7.3.3

**Depends** R (>= 4.1.0)

**Imports** checkCLI, dplyr, ggpp, purrr, tidyr, ggplot2, lemon, MASS, stats

**Suggests** knitr, rmarkdown, roxyglobals, tictoc

**VignetteBuilder** knitr

**Config/roxyglobals/filename** generated-globals.R

**Config/roxyglobals/unique** TRUE

**URL** https://luizesser.github.io/ECDFniche/

**NeedsCompilation** no

**Author** Luíz Fernando Esser [aut, cre, cph] (ORCID: <https://orcid.org/0000-0003-2982-7223>), Matheus Baumgartner [aut] (ORCID: <https://orcid.org/0000-0001-7472-8588>), Dayani Bailly [aut] (ORCID: <https://orcid.org/0000-0002-6954-9902>), Marcos R. Lima [aut] (ORCID: <https://orcid.org/0000-0002-5901-0911>), Reginaldo Ré [aut] (ORCID: <https://orcid.org/0000-0001-6452-3466>)

**Maintainer** Luíz Fernando Esser <luizesser@gmail.com>

**Repository** CRAN

**Date/Publication** 2026-04-27 18:40:09 UTC

# Contents



---

`create_distance_suitability_plot`
*Create distance–suitability plot*

---

## Description

Create distance–suitability plot

## Usage

```
create_distance_suitability_plot(analysis_results)
```

## Arguments

`analysis_results`
List returned by `ecdf_theoretical_niche()`.

## Value

A ggplot object.

## Examples

```
# Create ECDF-niche based on personalized options:
res <- ecdf_theoretical_niche(n = 3,
                              n_population = 20000,
                              sample_sizes = seq(50, 1000, 50),
                              seed = 123)

# Plot analysis results
create_distance_suitability_plot(res)
```

`ecdf_compare_niche` *Simulations and analyses of Mahalanobis distance-based habitat suitability*

## Description

The objective is to compare the performance of habitat suitability calculated based on chi-squared cumulative distribution function and Empirical Cumulative Distribution Function (ECDF)

## Usage

```
ecdf_compare_niche(
  p_vals = 1:5,
  n_vals = seq(20L, 500L, 20L),
  n_reps = 30L,
  seed = NULL
)
```

## Arguments

| | |
|---|---|
| `p_vals` | Integer vector; number of predictor variables (dimensions). |
| `n_vals` | Integer vector; number of records (sample sizes). |
| `n_reps` | Integer; number of replicates per combination. |
| `seed` | Optional integer for reproducibility. |

## Details

Performs replicated simulations of multivariate normal data to evaluate the agreement between suitability derived from chi-squared distribution and empirical cumulative distribution function (ECDF).

## Value

A list with:

- cor_plot: ggplot of correlation vs sample size.
- suit_plot: ggplot of suitability vs Mahalanobis distance.
- cond_plot: ggplot of correlation vs condition number.
- cor_df: raw correlation data.
- obs_df: observation-level data.
- cov_df: covariance diagnostics.

## Author(s)

Matheus T. Baumgartner

## Examples

```
# Create ECDF-niche based on personalized options:
n <- ecdf_compare_niche(p_vals = 1:3,
                        n_vals = seq(50L, 500L, 50L),
                        n_reps = 10L,
                        seed = 1991)
```

---

`ecdf_nonnormal_niche` *Compare ECDF and Chi-square suitability under non-normal data*

---

## Description

Script to run a simulation study to compare Chi-square vs. ECDF approaches to quantify habitat suitability based on bivariate non-normal data. Bivariate data was simulated based on environmental variables (temperature and precipitation) using Gaussian copulas. Temperature followed a normal distribution while precipitation followed a Weibull distribution. The choices of the distributions were based on Haddad (2021) - Theoretical and Applied Climatology (for temperature) and on the estimation of rainfall in milimeters by Wilks (1989) - Journal of Applied Meteorology. Because the relationship between temperature and precipitation is complex across space (Rodrigo, 2022 - Theoretical and Applied Climatology), we defined five correlation values between the two variables.

`temp_parameters` and `prec_parameters` must comply to `stats::qnorm` or `stats::qweibull`, depending on the function chosen on `temp_function` and `prec_function`. For `"qnorm"`, user can specify `mean` and `sd`, while for `"qweibull"`

## Usage

```
ecdf_nonnormal_niche(
  rho_vals = c(-0.7, -0.3, 0, 0.3, 0.7),
  n_vals = c(20L, 50L, 100L, 200L, 500L),
  n_reps = 10L,
  N_ref = 1e+05,
  temp_function = "qnorm",
  temp_parameters = list(mean = 20, sd = 5),
  prec_function = "qweibull",
  prec_parameters = list(shape = 2, scale = 10),
  seed = NULL
)
```

## Arguments

| | |
|---|---|
| `rho_vals` | Numeric vector; correlations between variables. |
| `n_vals` | Integer vector; sample sizes. |
| `n_reps` | Integer; number of replicates. |
| `N_ref` | Integer; size of reference population for "true" parameters. |

| | |
|---|---|
| temp_function | Character; function used to model temperature values. One of: "qnorm" or "qweibull". |
| temp_parameters | List; list organizing parameters to pass to temp_function. |
| prec_function | Character; function used to model precipitation values. One of: "qnorm" or "qweibull". |
| prec_parameters | List; list organizing parameters to pass to temp_function. |
| seed | Optional integer for reproducibility. |

## Details

Simulates bivariate environmental data using Gaussian copulas with non-normal marginals (Normal for temperature and Weibull for precipitation), and evaluates agreement between chi-squared and ECDF suitability.

## Value

A list with:

- suit_plot: ggplot of suitability vs Mahalanobis distance
- cor_df: correlation results
- obs_df: observation-level data

## Author(s)

Matheus T. Baumgartner

## Examples

```
# Create ECDF-niche based on personalized options:
n <- ecdf_nonnormal_niche(rho_vals = c(-0.7, -0.3, 0, 0.3, 0.7),
                          n_vals   = c(20L, 50L, 100L, 200L, 500L),
                          n_reps   = 10L,
                          N_ref    = 1e5,
                          seed     = 1991)
```

---

# ecdf_theoretical_niche

*Niche analysis using ECDF and chi-squared*

---

## Description

Simulate niche suitability from Mahalanobis distance using both chi-squared and empirical CDF transformations, for a given number of predictor variables.

## Usage

```
ecdf_theoretical_niche(
  n,
  n_population = 10000L,
  sample_sizes = seq(20L, 500L, 20L),
  seed = NULL
)
```

## Arguments

| | |
|---|---|
| n | Integer; number of predictor variables (dimensions). |
| n_population | Integer; size of simulated environmental population. |
| sample_sizes | Integer vector of sample sizes to evaluate. |
| seed | Optional integer seed for reproducibility. |

## Value

A list with:

- corplot: ggplot object with correlation vs sample size.
- sample_data: matrix of simulated sample points.
- sample_niche: numeric vector of "true" niche suitability.
- chisq_suits: numeric vector, 1 - pchisq(Mahalanobis).
- ecdf_suits: numeric vector, 1 - ECDF(Mahalanobis).
- mahal_dists: numeric vector of Mahalanobis distances.

## Author(s)

Luíz Fernando Esser

## Examples

```
# Create ECDF-niche based on personalized options:
n <- ecdf_theoretical_niche(n = 3,
                            n_population = 20000,
                            sample_sizes = seq(50, 1000, 50),
                            seed = 123)
```

`mahal.dist` *Mahalanobis Distance Classifier for Ecological Niche Modeling*

## Description

A custom `caret` model specification implementing a Mahalanobis distance-based classifier for ecological niche modeling (ENM) and species distribution modeling (SDM). This implementation supports both parametric (chi-squared) and nonparametric (empirical cumulative distribution function; ECDF) transformations of Mahalanobis distances into suitability scores.

## Usage

```
mahal.dist
```

## Format

An object of class `list` of length 12.

## Details

The model is trained using presence-only data to estimate the centroid and covariance structure of environmental conditions associated with species occurrences. Suitability is then derived as the inverse tail probability of the Mahalanobis distance between new observations and the estimated niche centroid.

Two approaches are available to transform Mahalanobis distances into probabilities:

- `"chisq"`: assumes distances follow a chi-squared distribution with degrees of freedom equal to the number of predictors.
- `"ecdf"`: uses the empirical cumulative distribution function of training distances, providing a nonparametric estimate of suitability.

The ECDF-based approach is particularly useful when the assumption of multivariate normality is violated, which is common in ecological data.

This model can be used within the `caret::train()` framework, enabling resampling, tuning, and ensemble modeling workflows for ecological niche modeling.

## Model Parameters

**abs** Logical. If `TRUE`, predictions are binarized using a fixed threshold (default: 0.05). If `FALSE`, the class with the highest predicted probability is returned.

**method** Character. Method used to convert Mahalanobis distances into suitability values. Options are `"chisq"` or `"ecdf"`.

## Details

The Mahalanobis distance defines an ellipsoidal niche in environmental space. Under the chi-squared formulation, suitability decreases as the distance from the niche centroid increases. The ECDF formulation relaxes distributional assumptions by estimating suitability directly from the empirical distribution of distances observed in presence data.

Predictions return class probabilities for `"presence"` and `"pseudoabsence"`, allowing flexible thresholding and ensemble integration.

## Usage in caret

This object can be supplied to `caret::train()` as a custom model:

```
library(caret)

model <- train(
  x = predictors,
  y = response,
  method = mahal.dist,
  trControl = trainControl(classProbs = TRUE)
)
```

You can also run only ECDF by adjusting the tuning grid:

```
library(caret)

grid <- expand.grid(
  abs = c(TRUE, FALSE),
  method = "ecdf"
)

model <- train(
  x = predictors,
  y = response,
  method = mahal.dist,
  tuneGrid = grid,
  trControl = trainControl(classProbs = TRUE)
)
```

## See Also

`mahalanobis`, `ecdf`, `train`

`run_ecdf_mahal_analysis` *Run full ECDF–Mahalanobis analysis*

## Description

Convenience function that reproduces the three figures from the original manuscript for 1–5 dimensions.

## Usage

```
run_ecdf_mahal_analysis(dims = 1:5, seed = 3L)
```

## Arguments

| | |
|---|---|
| `dims` | Integer vector of dimensions (default 1:5). |
| `seed` | Optional seed for reproducibility. |

## Value

A list containing:

- analyses: list of ecdf_theoretical_niche() outputs.
- figure1, figure2, figure3: grobs with arranged plots.

## Examples

```
# Recreate original manuscript output:
set.seed(3)
full_res <- run_ecdf_mahal_analysis(dims = 1:5)
```

# Index